

SGD: X - train set.

$$\begin{aligned} W_0 &\sim N(\dots) \\ \text{mini batch. } X &\sim 2^{\text{batch-size.}} \\ \underbrace{X_i}_{\text{mini batch.}} &\sim 2^{\text{batch-size.}} \\ W_{i+1} &= W_i - \gamma \cdot \nabla_w L(W_i, X_i) \end{aligned}$$

GD

$$\begin{aligned} X_i &\in 2^{\text{batch-size.}} \\ X_i &\subset \mathbb{R}^{128} \end{aligned}$$

$$z \in \mathbb{R}^{128}$$

$$h_1 = \psi_1(W_1^T \cdot z^{128})$$

$$W \in \mathbb{R}^{128 \times \mathbb{R}^{64}}$$

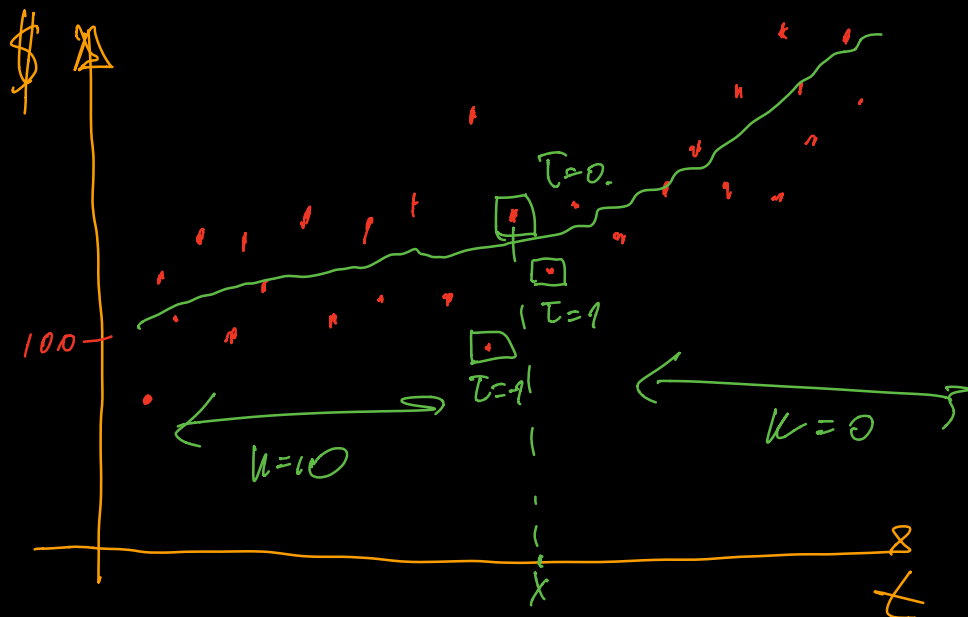
$$h_2 = \psi_2(W_2^T \cdot h_1)$$

$$h_i \in \mathbb{R}^{64}$$

\vdots

$$h_n = \psi_n(W_n h_{n-1})$$

MLP.



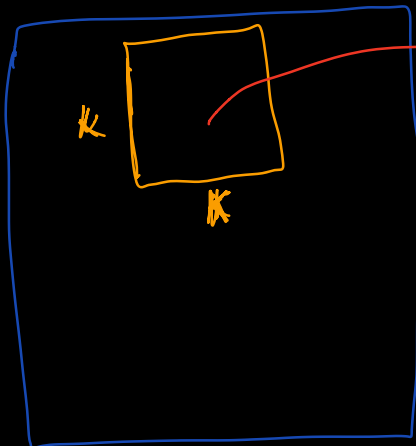
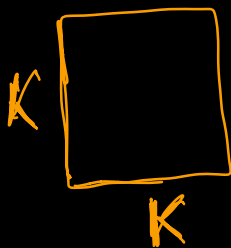
$$f: t \rightarrow \$$$

$$\underline{k: t \rightarrow k} \quad k \in \mathbb{R},$$

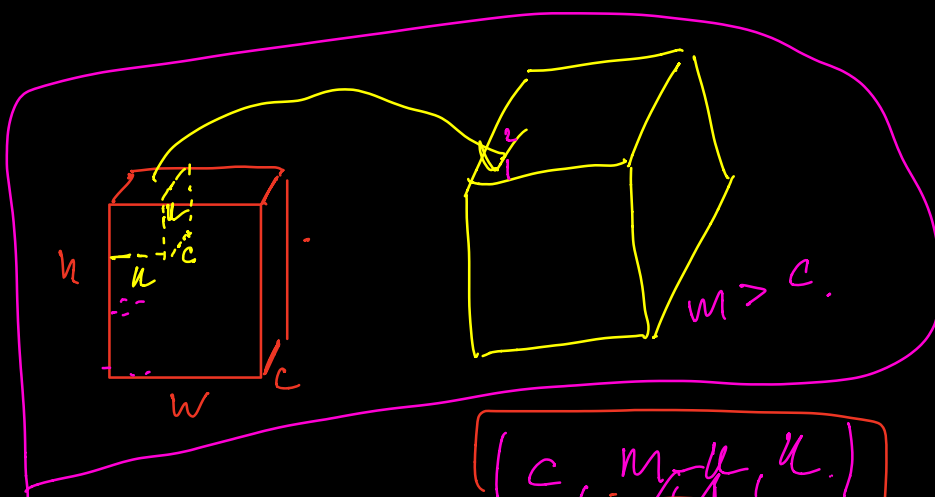
$$T = 10.$$

$$(f \times k)(x), \quad k =$$

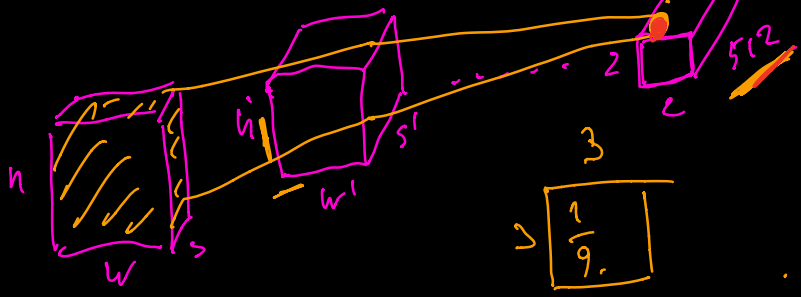
$$k = \begin{cases} T < -1, & 0 \\ T \in [-1, 1], & \frac{1}{3} \\ T > 1, & 0 \end{cases}$$



c, c_2, c_1



(c, m, u, u)



$$\frac{1}{9}$$